

Bài thực hành 9. TRIỂN KHAI ỨNG DỤNG HỌC MÁY

1. XÂY DỰNG ỨNG DỤNG WEB ĐƠN GIẢN VỚI STREAMLIT

- Website Streamlit: <https://streamlit.io/>.
- Tài liệu tham khảo: <https://docs.streamlit.io/>.
- Ví dụ minh họa:
 - <https://streamlit.io/components>.
 - <https://streamlit.io/gallery>.

2. BÀI TẬP

Yêu cầu

Tùy vào câu hỏi mà sinh viên xây dựng ứng dụng web đơn giản, có các thành phần giao diện và chức năng cần thiết để đáp ứng yêu cầu của câu hỏi.

Bài tập bắt buộc

- **Bài tập 1.** Xây dựng ứng dụng **khám phá dữ liệu**.
 - Các chức năng tối thiểu cần có: Upload file dữ liệu, hiển thị dữ liệu, thống kê mô tả dữ liệu, mô tả các thuộc tính, vẽ biểu đồ histogram biểu diễn sự phân bố giá trị của các thuộc tính, tính hệ số tương quan giữa các thuộc tính, chọn biến phụ thuộc và vẽ biểu đồ phân tán biểu diễn mối liên hệ giữa biến phụ thuộc và từng biến độc lập,...
 - Dữ liệu thử nghiệm: <https://www.kaggle.com/datasets/anikchandra70/petrol-consumption-multivariable-linear-regression>.
- **Bài tập 2.** Xây dựng ứng dụng nhận diện **bệnh nhiễm khuẩn phổi từ ảnh X-Quang**.
 - Nguồn dữ liệu: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>.
 - Input: Một tấm ảnh X-quang.
 - Output: Nhận dự đoán của tấm ảnh đó, gồm 1 trong 2 nhãn: NORMAL - bình thường và PNEUMONIA - bệnh.
- **Bài tập 3.** Xây dựng ứng dụng nhận dạng **cảm xúc từ câu phản hồi bằng văn bản của người dùng**.
 - Nguồn dữ liệu:
https://drive.google.com/drive/folders/1xclbjHHK58zk2X6iqbvMPS2rcy9y9E0X?usp=drive_open.
 - Input: Câu bình luận của người dùng.

- Output: Nhãn cảm xúc, gồm 1 trong 3 nhãn: tích cực, tiêu cực và trung tính.
- **Bài tập 4.** Xây dựng ứng dụng phân loại **loài hoa Iris dựa trên kích thước cánh hoa và đài hoa**.
 - Nguồn dữ liệu: Được cung cấp sẵn trong thư viện Sklearn.
 - Input: Các thông số về kích thước cánh hoa và đài hoa.
 - Output: Tên loài hoa, gồm 1 trong 3 loài: Setosa, Versicolor và Virginica.
- **Bài tập 5.** Xây dựng ứng dụng nhận dạng **ảnh chữ số viết tay**.
 - Nguồn dữ liệu: Được cung cấp sẵn trong thư viện Sklearn.
 - Input: Ảnh chụp chữ số viết tay.
 - Output: Nhãn chữ số, bao gồm các chữ số từ 0 đến 10.
- **Bài tập 6.** Xây dựng ứng dụng dự đoán **giới tính của người Việt Nam dựa trên họ tên**.
 - Nguồn dữ liệu: <https://github.com/JkUndead/UIT-ViNames-Dataset>.
 - Input: Họ tên người Việt Nam.
 - Output: Giới tính, gồm 1 trong 2 nhãn: 0 - nữ, 1 - nam.
- **Bài tập 7.** Xây dựng ứng dụng dự báo **tỷ lệ trúng tuyển vào trường cao đẳng tại Mỹ**.
 - Nguồn dữ liệu:
<https://www.kaggle.com/datasets/vishalkumbhar1997/admission-prediction-data>.
 - Input: Các thông số về thí sinh và trường cao đẳng.
 - Output: Xác suất thí sinh trúng tuyển vào trường cao đẳng.

Bài tập tự chọn

- **Bài tập 8.** Xây dựng ứng dụng nhận diện **món ăn Việt Nam**.
 - Nguồn dữ liệu:
<https://www.kaggle.com/datasets/quandang/vietnamese-foods>.
 - Input: Một tấm ảnh chụp món ăn.
 - Output: Tên của món ăn.
- **Bài tập 9.** Xây dựng ứng dụng nhận diện **rắn có độc hay không có độc**.
 - Nguồn dữ liệu: <https://github.com/arjun921/Indian-Snakes-Dataset>.
 - Input: Ảnh chụp con rắn.
 - Output: Nhãn nhận biết độc/không độc, gồm 1 trong 2 nhãn: 0 - không có độc, 1 - có độc.
- **Bài tập 10.** Xây dựng ứng dụng dự đoán **điểm đánh giá ứng dụng**

Instagram trên Play Store.

- Nguồn dữ liệu: <https://www.kaggle.com/datasets/saloni1712/instagram-play-store-reviews>.
- Input: Câu đánh giá về ứng dụng.
- Output: Điểm đánh giá của ứng dụng.

• Bài tập 11. Xây dựng ứng dụng nhận diện biển báo giao thông.

- Nguồn dữ liệu:
<https://www.kaggle.com/datasets/ahemateja19bec1025/traffic-sign-dataset-classification>.
- Input: Ảnh chụp biển báo giao thông.
- Output: Tên của biển báo giao thông.

• Bài tập 12. Xây dựng ứng dụng đánh giá chất lượng rượu vang.

- Nguồn dữ liệu: <https://www.kaggle.com/datasets/nareshbhat/wine-quality-binary-classification>.
- Input: Các thông số của rượu vang.
- Output: Nhãn chất lượng của rượu, gồm 1 trong 2 nhãn: good, bad.

• Bài tập 13. Xây dựng ứng dụng dự đoán giới tính dựa trên ảnh chụp.

- Nguồn dữ liệu: <https://www.kaggle.com/datasets/ashishjangra27/gender-detection-20k-images-celeba>.
- Input: Ảnh chụp người.
- Output: Nhãn giới tính, gồm 1 trong 2 nhãn: female, male.

• Bài tập 14. Xây dựng ứng dụng dự đoán giá kim cương.

- Nguồn dữ liệu:
<https://www.kaggle.com/datasets/shubhankitsirvaiya06/diamond-price-prediction>.
- Input: Các thông số về viên kim cương.
- Output: Giá của viên kim cương.